

**What is claimed is:**

SUB B17

1. A method of evaluating characters in a message to generate a search index, comprising the steps of:

- a) accepting an input of the characters of the message;
- b) evaluating the message by comparing the characters of the message to a predetermined set of candidate character sets to determine a match between the predetermined set of candidate character sets and the message; and
- c) generating a search index based on the results of the evaluation of the message and candidate character sets.

SUB C47

2. The method of claim 1, wherein the comparing of step (b) comprises the step of comparing each character of the message to an entry for each of the candidate character sets in a character table bank.

3. The method of claim 2, wherein the step of comparing each character comprises the step of testing the ability of each candidate character set to express that character by performing a logical mask between a universal code for that character and an indicator in the character table bank indicating whether each of the candidate character sets contains that character.

4. The method of claim 3, wherein the universal code is Unicode.

5. The method of claim 1, further comprising the step of (d) computing a total number of characters matched to each of the candidate character sets.

6. The method of claim 1, further comprising the step of (e) selecting a best match based upon the total number of characters matched to each of the candidate character sets.

7. The method of claim 1, further comprising the step of (f) evaluating the characters of a query string.

8. The method of claim 7, further comprising the step of (g) performing a search of the query string against search indices whose character set matches the character set of the query string.

SUB B2

9. A system for evaluating characters in a message to generate a search index, comprising:

an input interface to accept an input of the characters of the message; and  
a processor unit, connected to the input interface, the processor unit evaluating the message by comparing the characters of the message to a predetermined set of candidate character sets to determine a match between the predetermined set of candidate character sets and the message, and generating a search index based on the results of the evaluation of the message and candidate character sets.

SUB C6

10. The system of claim 9, wherein the processor unit compares each character of the message to an entry for each of the candidate character sets in a character table bank.

11. The system of claim 10, wherein the processor unit tests the ability of each candidate character set to express that character by performing a logical mask between a universal code for that character and an indicator in the character table bank indicating whether each of the candidate character sets contains that character.

12. The system of claim 11, wherein the universal code is Unicode.

13. The system of claim 9, wherein the processor unit computes a total number of characters matched to each of the candidate character sets.

14. The system of claim 9, wherein the processor unit selects a best match based upon the total number of characters matched to each of the candidate character sets.

15. The system of claim 9, wherein the processor unit evaluates the characters of a query string.

16. The system of claim 15, wherein the processor unit performs a search of the query string against search indices whose character set matches the character set of the query string.

SUB B3

17. A system for evaluating characters in a message to generate a search index, comprising:

input interface means to accept an input of the characters of the message; and  
processor means, connected to the input interface means, the processor means evaluating the message by comparing the characters of the message to a

~~predetermined set of candidate character sets to determine a match between the predetermined set of candidate character sets and the message, and generating a search index based on the results of the evaluation of the message and candidate character sets.~~

SUB C87

~~18. The system of claim 17, wherein the processor means compares each character of the message to an entry for each of the candidate character sets in a character table bank.~~

19. The system of claim 18, wherein the processor means tests the ability of each candidate character set to express that character by performing a logical mask between a universal code for that character and an indicator in the character table bank indicating whether each of the candidate character sets contains that character.

20. The system of claim 19, wherein the universal code is Unicode.

21. The system of claim 17, wherein the processor means computes a total number of characters matched to each of the candidate character sets.

22. The system of claim 17, wherein the processor means selects a best match based upon the total number of characters matched to each of the candidate character sets.

23. The system of claim 17, wherein the processor means evaluates the characters of a query string.

24. The system of claim 23, wherein the processor means performs a search of the query string against search indices whose character set matches the character set of the query string.

SUB B47

~~25. A storage medium for storing machine readable code, the machine readable code being executable to evaluate characters in an electronic message according to the steps of:~~

- ~~a) accepting an input of the characters of the message;~~
- ~~b) evaluating the message by comparing the characters of the message to a predetermined set of candidate character sets to determine a match between the predetermined set of candidate character sets and the message; and~~

~~c)~~ generating a search index based on the results of the evaluation of the message and the candidate character sets.

SUB C10>

26. The storage medium of claim 25, wherein the comparing of step (b) comprises the step of comparing each character of the message to an entry to each of the candidate character sets in a character table bank.

27. The storage medium of claim 26, wherein the step of comparing each character comprises the step of testing the ability of each candidate character set to express that character by performing a logical mask between a universal code for that character and an indicator in the character table bank indicating whether each of the candidate character sets contains that character.

28. The storage medium of claim 27, wherein the universal code is Unicode.

29. The storage medium of claim 25, wherein the steps further comprise the step of (d) computing a total number of characters matched to each of the candidate character sets.

30. The storage medium of claim 25, wherein the steps further comprise the step of (e) selecting a best match based upon the total number of characters matched to each of the candidate character sets.

31. The storage medium of claim 25, wherein the steps further comprise the step of (f) evaluating the characters of a query string.

32. The storage medium of claim 31, wherein the steps further comprise the step of (g) performing a search of the query string against search indices whose character set matches the character set of the query string.

ADD A2>

ADD B6>

ADD C12>